



日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 4 年 1 月 2 9 日
Date of Application:

出 願 番 号 特 願 2 0 0 4 - 0 2 1 2 0 8
Application Number:
[ST. 10/C] : [J P 2 0 0 4 - 0 2 1 2 0 8]

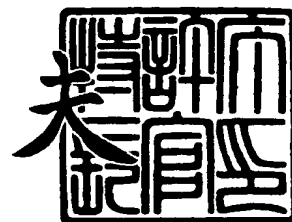
出 願 人 株式会社日立製作所
Applicant(s):

U.S. Appln. Filed 3-22-04
Inventor: Y. Nonaka et al
Mattingly Stanger & Malur
Docket n17-417

2 0 0 4 年 3 月 8 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



出証番号 出証特 2 0 0 4 - 3 0 1 7 8 1 9



【書類名】 特許願
【整理番号】 NT03P0793
【提出日】 平成16年 1月29日
【あて先】 特許庁長官 殿
【国際特許分類】 G06F 12/00514
G06F 13/12

【発明者】
【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
システム開発研究所内
【氏名】 野中 裕介

【発明者】
【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
システム開発研究所内
【氏名】 松並 直人

【発明者】
【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
システム開発研究所内
【氏名】 八木沢 育哉

【発明者】
【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
システム開発研究所内
【氏名】 西本 哲

【特許出願人】
【識別番号】 000005108
【氏名又は名称】 株式会社日立製作所

【代理人】
【識別番号】 100068504
【弁理士】
【氏名又は名称】 小川 勝男
【電話番号】 03-3537-1621

【選任した代理人】
【識別番号】 100086656
【弁理士】
【氏名又は名称】 田中 恭助
【電話番号】 03-3537-1621

【選任した代理人】
【識別番号】 100094352
【弁理士】
【氏名又は名称】 佐々木 孝
【電話番号】 03-3537-1621

【手数料の表示】
【予納台帳番号】 081423
【納付金額】 21,000円

【提出物件の目録】
【物件名】 特許請求の範囲 1
【物件名】 明細書 1
【物件名】 図面 1
【物件名】 要約書 1

【書類名】 特許請求の範囲**【請求項 1】**

ブロック単位の I/O コマンドによる SAN インタフェースと、ファイル単位 of I/O コマンドによる NAS インタフェースとによりデータのアクセスをおこなう記憶装置において、

ディスク装置と、ディスク装置を制御するコントローラとを有し、

前記コントローラは、

計算機からネットワークを介して前記ファイル単位 of I/O コマンドを受け付ける NAS 制御ボードと、

計算機からネットワークを介して前記ブロック単位 of I/O コマンドを受け付け、それにより前記ディスク装置に記憶されたデータをアクセスする SAN 制御ボードとを有し、

前記 NAS 制御ボードは、前記ファイル単位 of I/O コマンドを、ブロック単位 of I/O コマンドに変換して、前記 SAN 制御ボードに送信し、

前記 SAN 制御ボードは、前記 NAS 制御ボードから送信されるブロック単位 of I/O コマンドにより、前記ディスク装置に記憶されたデータをアクセスすることを特徴とする記憶装置。

【請求項 2】

前記 NAS 制御ボード上のプロセッサにより動作するコマンド変換プログラムにより、前記ファイル単位 of I/O コマンドを、前記ブロック単位 of I/O コマンドに変換することを特徴とする請求項 1 記載の記憶装置。

【請求項 3】

前記 SAN 制御ボード上のプロセッサにより動作するディスク装置制御プログラムは、前記 NAS 制御ボードから送信されるファイル単位 of I/O コマンドの有無の如何によらず動作することを特徴とする請求項 1 記載の記憶装置。

【請求項 4】

前記 NAS 制御ボード上のプロセッサと、前記 SAN 制御ボード上のプロセッサとは、両者を結合する内部バスにより通信することを特徴とする請求項 1 記載の記憶装置。

【請求項 5】

前記 NAS 制御ボードと、前記 SAN 制御ボードとは、別々のメモリを有することを特徴とする請求項 1 記載の記憶装置。

【請求項 6】

前記 NAS 制御ボードは、前記 SAN 制御ボードに装着され、

前記 SAN 制御ボードの動作中に NAS 制御ボードが着脱されても、この記憶装置がネットワークを介して計算機から受け付けたブロック単位 of I/O コマンドを、前記 SAN 制御ボードが処理することが可能なことを特徴とする請求項 1 記載の記憶装置。

【請求項 7】

前記ブロック単位 of I/O コマンドと前記ファイル単位 of I/O コマンドとのいずれかを優先的に処理してデータをアクセスすることが可能なことを特徴とする請求項 1 記載の記憶装置。

【請求項 8】

ブロック単位 of I/O コマンドによる SAN インタフェースと、ファイル単位 of I/O コマンドによる NAS インタフェースとによりデータのアクセスをおこなう記憶装置において、

この記憶装置は、ディスク装置を制御するコントローラを有し、

前記コントローラは、

前記ファイル単位 of I/O コマンドを受け付けるための NAS 制御ボードと、

前記ブロック単位 of I/O コマンドを受け付けるための SAN 制御ボードとを備え、

前記 NAS 制御ボードは、前記 NAS インタフェースの通信コントローラと、NAS プロセッサと、NAS メモリと、前記 NAS メモリのコントローラとを有し、

前記 SAN 制御ボードは、前記 SAN インタフェースの通信コントローラと、SAN プ

ロセッサと、SANメモリと、前記SANメモリのコントローラと、前記NASプロセッサと前記SANプロセッサのプロセッサ間通信コントローラと、ディスク装置を制御するディスクコントローラとを有し、

前記NAS制御ボードは、前記SAN制御ボードに装着され、

前記SAN制御ボードの動作中にNAS制御ボードが着脱されても、前記SANインタフェースの通信コントローラを介して受信したブロック単位のI/Oコマンドを処理することが可能であって、

前記NASメモリには、コマンド変換プログラムが格納され、

前記SANメモリには、ディスクアレ制御プログラムが格納され、

前記NASプロセッサは、前記コマンド変換プログラムを実行して、前記NASインタフェースの通信コントローラを介して受信したファイル単位のI/Oコマンドを、前記ブロック単位のI/Oコマンドに変換し、前記NASメモリ内の通信用領域に格納して、

前記プロセッサ間通信コントローラは、前記NASメモリ内の通信用領域に格納された前記ブロック単位のI/Oコマンドを、前記SANメモリ内の通信用領域に転送して、

前記SANプロセッサは、前記ディスクアレ制御プログラムを実行して、前記SANメモリ内の通信用領域に転送された前記ブロック単位のI/Oコマンドと、前記SANインタフェースの通信コントローラを介して受信した前記ブロック単位のI/Oコマンドをキューイングして、前記キューイングされたブロック単位のI/Oコマンドを参照して、前記ディスクコントローラを介して、この記憶装置に接続されたディスク装置に記憶されたデータをアクセスすることを特徴とする記憶装置。

【請求項9】

ブロック単位のI/OコマンドによるSANインタフェースと、ファイル単位のI/OコマンドによるNASインタフェースとによりデータのアクセスをおこなう記憶装置において、

この記憶装置は、ディスク装置を制御する制御ボードを有し、

前記制御ボードは、1ボードにより実装され、

前記ファイル単位のI/Oコマンドを、ブロック単位のI/Oコマンドに変換し、前記ブロック単位のI/Oコマンドにより、前記ディスク装置に記憶されたデータをアクセスすることを特徴とする記憶装置。

【請求項10】

前記ファイル単位のI/Oコマンドをブロック単位のI/Oコマンドに変換するプログラムと、前記ブロック単位のI/Oコマンドを処理するプログラムが独立して動作可能なことを特徴とする請求項9記載の記憶装置。

【請求項11】

前記ファイル単位のI/Oコマンドをブロック単位のI/Oコマンドに変換するプログラムと、前記ブロック単位のI/Oコマンドを処理するプログラムとが単一のメモリを共有することを特徴とする請求項9記載の記憶装置。

【請求項12】

ブロック単位のI/OコマンドによるSANインタフェースと、ファイル単位のI/OコマンドによるNASインタフェースとによりデータのアクセスをおこなう記憶装置の制御方法において、

記憶装置は、

ネットワークを介して計算機から前記ファイル単位のI/Oコマンドを受け付けるNAS制御ボードと、

ネットワークを介して計算機から前記ブロック単位のI/Oコマンドを受け付け、それにより前記ディスク装置に記憶されたデータをアクセスするSAN制御ボードとを有し、

前記ファイル単位のI/Oコマンドを、ブロック単位のI/Oコマンドに変換するステップと、

前記変換された前記ブロック単位のI/OコマンドをSAN制御ボードに送信するステップと、

前記 N A S 制御ボードから送信されるブロック単位の I / O コマンドにより、前記ディスク装置に記憶されたデータをアクセスするステップとを有することを特徴とする記憶装置の制御方法。

【請求項 13】

前記ファイル単位の I / O コマンドをブロック単位の I / O コマンドに変換するステップの処理は、前記コントローラ上のプロセッサで動作するコマンド変換プログラムによりおこなわれることを特徴とすること請求項 12 記載の記憶装置の制御方法。

【請求項 14】

前記ファイル単位の I / O コマンドを変換したブロック単位の I / O コマンドは、外部から受け付けたブロック単位の I / O コマンドと同一のコマンドキューに格納されて実行されることを特徴とする請求項 12 記載の記憶装置の制御方法。

【請求項 15】

前記ファイル単位の I / O コマンドを変換したブロック単位の I / O コマンドは、外部から受け付けたブロック単位の I / O コマンドと別のコマンドキューに格納されて実行されることを特徴とする請求項 12 記載の記憶装置の制御方法。

【請求項 16】

前記コントローラが、前記ファイル単位の I / O コマンドが変換されたブロック単位の I / O コマンドと、外部から受け付けたブロック単位の I / O コマンドのいずれかに優先度を付けて実行することが可能なことを特徴とする請求項 15 記載の記憶装置の制御方法。

【書類名】 明細書**【発明の名称】** 複数インタフェースを有する記憶装置、および、その記憶装置の制御方法**【技術分野】****【0001】**

本発明は、複数インタフェースを有する記憶装置に係り、特に、SANインタフェースとNASインタフェースのコマンドの両者によりデータアクセスをおこなうことのできるハイブリッド型の記憶装置に関する。

【背景技術】**【0002】**

近年、情報システムの適用分野の拡大および計算機システムの分散化の進行にともない、一つのサイト内で管理されるホストの数が飛躍的に増大している。これにともない、計算機システムの管理コストの増大が問題となっている。特に、分散したホストが持つ記憶装置の管理を個別に行えば、運用に必要とされる容量を事前に予測することの難しさから、容量の追加をおこなう必要が生じた場合に各ホストに個別に記憶装置を増設する必要がある、管理コストは大きくなる。

【0003】

この問題を解決するための技術として、SAN(Storage Area Network)およびNAS(Network Attached Storage)が利用されている。いずれの技術も、分散したホストが個別に持つ記憶装置の集約を目的としているが、それぞれの技術が持つ特徴から異なる適用分野に用いられている。

【0004】

SANは、複数の記憶装置と複数のホストを、ホスト-記憶装置間通信専用のネットワークによって接続し、高速で遅延が少ないI/Oを実現する技術である。SANにおいては、ホストと記憶装置の間のI/Oはブロック単位でおこなわれる。ブロックとは、記憶装置の容量をあるサイズに分割した、アドレスによって識別される固定長のデータの管理単位である。ブロック単位のI/Oが適したアプリケーションとしては、主にデータベースが挙げられる。複数のデータベースサーバのボリュームを単一の記憶装置に集約し、管理を一元化することによって、計算機システムの管理コストを削減することが可能である。また、SANが高信頼な専用ネットワークであり、効率の良い通信がおこなえることから、ボリューム単位のデータ転送のための通信路としての使用も有効である。ボリューム単位のデータ転送は、バックアップや、災害対策のための、遠隔の記憶装置へボリューム複製の目的で利用される。

【0005】

一方、NASは、ファイルサーバ機能を持つ記憶装置であり、LAN(Local Area Network)上の複数のホストに対するファイルサービスを提供する。NASとホストの間のI/Oは、ファイル単位でおこなわれる。ファイルは、一般的に文字列の識別子によって識別される、可変長のデータの管理単位である。ファイル単位のI/Oが適したアプリケーションとしては、複数のホストによるファイル共有が挙げられる。複数のWebサーバによる同一コンテンツのサービスや、複数のオフィスPCによる文書の共有を実現することができる。

【0006】

このように、SANとNASは互いを補完する関係にある技術である。したがって、SANに対応した記憶装置とNASを単一の装置に統合することで、さらに管理コストが削減できるという発想に到る。下記非特許文献1および特許文献1に記載された技術は、記憶装置の制御プログラムをSANとNASの両方の制御をおこなえるようにすることで、SANとNASに両対応した記憶装置を実現するものである。この技術を使用した記憶装置は、SANとNASのインタフェースを持ち、記憶容量をSANとNASに自由に割り当てることができる。

【0007】**【特許文献1】** 特開2003-162439号公報

【0008】

【非特許文献1】Stephen Daniel, “Converging SAN and NAS Storage A Comparison of Unified and Gateway Solutions,” Network Appliance Inc. White Paper、OCT 2002.

【発明の開示】

【発明が解決しようとする課題】

【0009】

上記非特許文献1の従来技術により、SANとNASに単一の記憶装置で対応することができる。しかしながら、SANとNASを単一の制御プログラムによって実現するので、互いの機能が干渉するという問題がある。NAS機能の障害によってSAN機能が停止する可能性が高く、耐障害性に問題がある。また、NAS機能のみを停止して保守をおこなうことができない。また、NASの負荷が高くなると、SANの性能が著しく低下する。逆に、SANの負荷が高くなると、NASの性能が著しく低下する。このような現象は、システム全体の性能設計を困難にする。一般に、SANがストレージシステムの基幹となることが多いため、NAS機能が停止してもSAN機能による記憶装置のデータアクセスが止まらないことが要望される。

【0010】

一方、上記特許文献1に記述された従来技術により、NAS機能とSAN機能の独立性を高め、NAS機能の障害によってSAN機能が停止しない記憶装置を実現することができる。しかしながら、NAS機能およびSAN機能を持つ複数のインタフェースがそれぞれディスクの制御をおこなうため、内部ネットワークと共有メモリによって通信しなければならずハードウェアが高価になり、インタフェース数が小規模構成の記憶装置に適用するのは困難である。

【0011】

本発明は、上記従来技術の問題点を解決するためになされたもので、その目的は、SANインタフェースとNASインタフェースの両者を有するハイブリッド型の記憶装置において、NAS機能の障害やNASの負荷などにかかわらず、SAN機能として独立に動作し、しかも、簡易なハードウェアで実現する記憶装置を提供することにある。

【課題を解決するための手段】

【0012】

本発明では、ブロック単位のI/OコマンドによるSANインタフェースと、ファイル単位のI/OコマンドによるNASインタフェースとによりデータのアクセスをおこなう記憶装置において、ディスク装置を制御するコントローラを、ファイル単位のI/Oコマンドを受け付けるNAS制御ボードと、ブロック単位のI/Oコマンドを受け付け、それによりディスク装置に記憶されたデータをアクセスするSAN制御ボードとで構成する。

【0013】

そして、NAS制御ボードは、受け付けたファイル単位のI/Oコマンドを、ブロック単位のI/Oコマンドに変換して、SAN制御ボードに送信する。

【0014】

SAN制御ボードは、SANインタフェースから受け付けるブロック単位のI/Oコマンドと、NAS制御ボードから送信されるブロック単位のI/Oコマンドの両者により、ディスク装置に記憶されたデータをアクセスする。

【発明の効果】

【0015】

本発明によれば、SANインタフェースとNASインタフェースの両者を有するハイブリッド型の記憶装置において、NAS機能の障害やNASの負荷などにかかわらず、SAN機能として独立に動作し、しかも、簡易なハードウェアで実現することができる。

【発明を実施するための最良の形態】

【0016】

以下、本発明に係る各実施形態を、図1ないし図11を用いて説明する。

【0017】**〔実施形態1〕**

以下、本発明に係る第一の実施形態を、図1ないし図9を用いて説明する。

【0018】

先ず、図1を用いて本発明の第一の実施形態に係る記憶装置を用いた計算機システムの構成について説明する。

図1は、本発明の第一の実施形態に係る記憶装置を用いた計算機システムの構成図である。

【0019】

第一の実施形態における記憶装置100は、ファイバチャネルにより構成されたSAN50を経由して複数の計算機（以下、SANクライアント30と呼ぶ）に接続されており、SANクライアント30からのI/Oコマンドを受け付ける。また、LAN40を経由して複数の計算機（以下、NASクライアント20と呼ぶ）に接続されており、NASクライアント20からのI/Oコマンドを受け付ける。SAN50は、ファイバチャネル以外にも、iSCSIプロトコルが使用可能なIPネットワークの場合もある。

【0020】

記憶装置100は、ディスク装置105とコントローラ110によって構成されている。

【0021】

ディスク装置105は、複数のディスクドライブ150と、ディスクドライブ150間とコントローラ110とを接続する通信路155からなる。

【0022】

コントローラ110は、通信路155を経由して、任意のディスクドライブ150に対してI/Oをおこなうことができる。ディスクドライブ150と通信路155が、コントローラ110に含まれる形態もある。記憶装置100には、複数のディスク装置105が含まれる形態もある。

【0023】

コントローラ110は、SAN制御ボード140が主な構成部品であり、このSAN制御ボード上にNAS制御ボード145を実装するようになっている。SAN制御ボード140は、コントローラ110内に二枚搭載される形態もある。その形態は、一枚の障害時にもう一枚が処理を引き継ぐ冗長構成とすることで信頼化性を高めることを目的としている。

【0024】

ここで、複数のディスクドライブ150の記憶領域の一部を集めてグループ化して、連続したアドレス空間を割り当てた論理的な記憶領域を「論理ボリューム」といい、論理ボリュームのアドレス空間を使用するブロック単位のI/Oコマンドを「論理ブロックコマンド」という。これに対して、各ディスクドライブ150固有のアドレス空間を使用するブロック単位のI/Oコマンドを「物理ブロックコマンド」という。

【0025】

このSAN制御ボード140は、ディスク装置105、管理端末60、SAN50と接続されており、SANクライアント30から受信した論理ブロックコマンドに従って、ディスク装置105を構成するディスクドライブ150に対して物理ブロックコマンドを送信する機能を持つ。

【0026】

NAS制御ボード145は、SAN制御ボード140の上のコネクタ160に着脱141することができ、LAN40と接続されている。NAS制御ボード145は、NASクライアント20から受信したファイル単位のI/Oコマンド（以下、「ファイルコマンド」という）を論理ブロックコマンドに変換し、SAN制御ボード140に対して送信する機能を持つ。

【0027】

SAN制御ボード140が論理ボリュームとディスクの対応関係を記述した構成情報165を管理することによって、NAS制御ボード145が使用できるNAS領域170とSANクライアント30が使用できるSAN領域175にディスクドライブ150を自由に割り当てることができる。NAS制御ボード145と、SAN制御ボード140が、同一基板上に実装される形態もあるが、これは実施形態2で説明する。

【0028】

管理端末60は、LAN等によりSAN制御ボード140およびNAS制御ボード145と接続され、画面・キーボード・マウスを用いて、記憶装置100に対する管理操作をおこなう。管理操作には、記憶装置100の起動および停止、NAS制御ボード145の起動および停止、SAN制御ボード140とNAS制御ボード145を含む記憶装置100の設定変更および障害情報の参照が含まれる。

【0029】

次に、図2を用いてNAS制御ボードとSAN制御ボードの構成について説明する。

図2は、NAS制御ボードとSAN制御ボードの詳細なブロック図である。

【0030】

SANプロセッサ1409は、主にディスクアレイ制御プログラム1405とNAS制御ボード管理プログラム1406を解釈・実行するプロセッサであり、一時記憶装置としてSANメモリコントローラ1410に接続されたSANメモリ1407を使用する。SANメモリ1407に格納されているディスクアレイ制御プログラム1405は、ファイバチャネルコントローラ1401が受信した論理ブロックコマンドを解釈し、物理ブロックコマンドに変換することでディスクコントローラ1403を制御する機能を持つ。ディスクアレイ制御プログラム1405は、ディスクキャッシュ1402を用いて、コマンドに対する応答速度を向上させる。

【0031】

NAS制御ボード管理プログラム1406は、管理端末60からの管理操作に従って、NAS制御ボード145の起動および停止や、障害情報の参照をおこなう機能を持つ。これらのプログラムは、SANプロセッサ1409によって実行される。

【0032】

NAS制御ボード145上のNASプロセッサ1457は、主にコマンド変換プログラム1453を解釈・実行するプログラムであり、一時記憶装置としてNASメモリコントローラ1456に接続されたNASメモリ1452を使用する。NASメモリ1452に格納されているコマンド変換プログラム1453は、LANコントローラ1451が受信したファイルコマンドを解釈し、論理ブロックコマンドに変換する機能を持つ。コマンド変換プログラム1453は、作成した論理ブロックコマンドをSANプロセッサ1409に送信する。

【0033】

SANプロセッサ1409上で実行されるディスクアレイ制御プログラム1405は、NAS制御ボード145から送信されてきた論理ブロックコマンドを、物理ブロックコマンドに変換する機能を持つ。これは、ファイバチャネルコントローラ1401が受信した論理ブロックコマンドを物理ブロックコマンドに変換するときの処理と同様である。

【0034】

論理ブロックコマンドをNASプロセッサ1457からSANプロセッサ1409へ送信するときには、コマンド変換プログラム1453は、プロセッサ間通信コントローラ1408を制御する。論理ブロックコマンドは、通信用領域1455に格納され、例えば、プロセッサ間通信コントローラ1408が持つDMA機能により通信用領域1404に転送される。コマンドの実行結果であるステータスは、コマンド完了後、通信用領域1404から通信用領域1455に転送される。プロセッサ間通信コントローラ1408が、SANプロセッサ1409およびNASプロセッサ1457への割り込み機能を持っていれば、コマンドまたはステータスの送信完了を各プロセッサに即座に通知できる。

【0035】

NASクライアント20からのファイルコマンドが記憶装置への書き込みを要求するものであった場合には、データは、ファイルキャッシュ1454に一時的に格納される。次に、前記手順で論理ブロックコマンドが通信用領域1404に転送される。次に、ディスクアレイ制御プログラム1405が、転送された論理ブロックコマンドに基づいて、格納されているデータをディスクキャッシュ1402に転送する。

【0036】

また、NASクライアント20からのファイルコマンドが記憶装置からの読み込みを要求するものであった場合、前記手順で通信用領域1404に転送された論理ブロックコマンドに従ってディスクアレイ制御プログラム1405がディスクキャッシュ1402に読み込んだデータを、ファイルキャッシュ1454に転送する。NAS制御ボード145とSAN制御ボード140の間のコネクタ160に接続されたバスにより、論理ブロックコマンドやデータの送受信をおこなうことができ、また、SAN制御ボード140の通電中にも、NAS制御ボード145の着脱ができる。このような条件を満たすバスには、例えば、PCIバスがある。NAS制御ボード145がSAN制御ボード140上に複数設置される形態もある。このような形態では、プロセッサ間通信コントローラ1408がNAS制御ボード145の数だけ設置され、各NAS制御ボード145とSANプロセッサ1409との通信に使用される。プロセッサ間通信コントローラ1408がNAS制御ボード145上に実装される形態もある。

【0037】

上述のようにディスクアレイ制御プログラム1405は、SANを経由してクライアントから記憶装置にブロック単位で発行され、ディスクアレイ内の複数のディスクを用いて論理的に定義される論理ボリュームを対象に発行されるI/Oコマンドを、論理ボリュームを構成するディスクへのI/Oコマンドに変換する機能を持つプログラムである。

【0038】

また、コマンド変換プログラム1453は、LAN上のクライアントから発行されるファイル単位のI/Oコマンドを、ディスクアレイ制御プログラムによって処理可能なブロック単位のI/Oコマンドに変換する機能を持つプログラムである。

【0039】

ディスクアレイ制御プログラム1405は、コマンド変換プログラム1453の機能に依存せず単独で動作できる。したがって、プログラムの不具合等によりコマンド変換プログラムの処理が継続できなくなった場合でも、ディスクアレイ制御プログラム1405は処理を継続することができ、SAN上のクライアントからのI/Oコマンドの処理を続けることができる。

【0040】

これにより、NAS機能のみを停止して保守をおこなうことができる。また、ディスクアレイ制御プログラムが動作するプロセッサを複数にし、その一部のみがコマンド変換プログラム1453からのI/Oコマンドを処理するようにすることで、NASの負荷が高いときにも、SANの性能は一定以上に保たれる。また、SANの負荷が高いときにも、NASの性能は一定以上に保たれる。

【0041】

コマンド変換プログラム1453とディスクアレイ制御プログラム1405の通信には内部バスを用い、ディスクアレイ制御プログラム1405がディスク装置105の制御をおこなうことで内部ネットワークや共有メモリを不要とし、低コストでの実装を可能にする。

【0042】

次に、図3ないし図9を用いて本発明の第一の実施形態に係る記憶装置の制御動作について説明する。

まず、図3により図2のディスクアレイ制御プログラム1405が、図2のコマンド変換プログラム1453またはファイバチャネルコントローラ1401からのコマンドを受信した時に、コマンドキューに格納するまでの処理内容について説明する。

図3は、コマンド変換処理の流れを示すフローチャートである。

【0043】

コマンド変換プログラム1453は、コマンド変換（ステップ300）をおこなった後に、そのコマンドを通信用領域1404に転送する（ステップ302）。その後、プロセッサ間通信コントローラ1408のレジスタを設定することにより（ステップ305）、コマンド受信待ち（ステップ310）状態にあるディスクアレイ制御プログラム1405は、割り込みを受信する。このとき、コマンドキューイング処理中かどうかの判定をおこない（ステップ315）、処理中でなければ、受信したコマンドをキューイングし（ステップ320）、処理中であれば、終了するまで待つ。コマンドキューイング処理中である場合、ファイバチャネルコントローラ1401が受信したコマンドのキューイングを実行中である。そのため、ステップ315の判定処理により、コマンド変換プログラム1453とファイバチャネルコントローラ1401から同時にコマンドを受信したときの排他制御を実現する。この排他制御は、例えば、プロセッサ間通信コントローラ1408とファイバチャネルコントローラ1401がSANプロセッサ1409に対して送信する割り込みの優先度を同じにし、割り込み処理としてコマンドキューイングを実行することにより実現できる。

【0044】

次に、図4および図5によりコマンドキューイングの際のコマンドのキューデータ構造について説明する。

図4および図5は、コマンドのキューのデータ構造の例を示す図である。

【0045】

これは、図2のコマンド変換プログラム1453により変換されたコマンドと、ファイバチャネルコントローラ1401から送られてきたコマンドの両者を保存するコマンドキューのデータ構造を示すものである。

【0046】

図4に示されるコマンドキューのデータ構造の例では、コマンドキュー先頭ポインタ400は、次に実行されるコマンドのアドレスを指し、コマンドキュー最後尾ポインタ410は、新たに受信したコマンドを後続に追加するコマンドを指す。コマンドキューには、コマンド変換プログラム1453から受信したNASコマンド415と、ファイバチャネルコントローラ1401から受信したSANコマンド420が混在しており、このコマンドキューによりコマンドを処理した場合には、全てのコマンドは受信順に逐次処理される。

【0047】

なお、コマンドキューは、図2のSANメモリ1407の中に保持される。

【0048】

図5に示されるコマンドキューのデータ構造は、図4とは異なったものであり、NASコマンド515と、SANコマンド565を別のキューで保存している。

【0049】

コマンド変換プログラム1453から受信したコマンドは、NASコマンドキュー最後尾ポインタ510の指すコマンドの後続に追加し、ファイバチャネルコントローラ1401から受信したコマンドは、SANコマンドキュー最後尾ポインタ560の指すコマンドの後続に追加する。

【0050】

このコマンドキューで特徴的なことは、NASコマンドとSANコマンドをどちらを優先的に処理するかを制御できることである。すなわち、次に実行するコマンドを、NASコマンドキュー先頭ポインタ500から先に検索すれば、図1のNASクライアント20からのI/Oが優先的に処理され、SANコマンドキュー先頭ポインタ550から先に検索すれば、図1のSANクライアント30からのI/Oが優先的に処理されることになる。

【0051】

次に、図6により本発明の第一の実施形態に係るコマンド変換プログラムとディスクアレイ制御プログラムの起動処理について説明する。

図6は、本発明の第一の実施形態に係るコマンド変換プログラムとディスクアレイ制御プログラムの起動するときの処理の流れを示すフローチャートである。

【0052】

図1の記憶装置100の電源投入後、SAN制御ボード140とNAS制御ボード145が初期化される。

【0053】

SAN制御ボード140の初期化後、IPLにより図2のSANメモリ1407にディスクアレイ制御プログラム1405とNAS制御ボード管理プログラム1406がディスクドライブ150よりロードされる。

【0054】

SANプロセッサ1409は、ディスクアレイ制御プログラム1405の実行を開始する。NAS制御ボード145の初期化後、IPLによりコマンド変換プログラム1453の起動プログラムがNAS制御ボード145上のROMからロードされる。コマンド変換プログラム1453は、ディスクドライブ150に保存されているが、IPLはディスクドライブ150にアクセスできないため、先にコマンド変換プログラム1453の起動プログラムがロードされるようにしたものである。この起動プログラムは、ディスクアレイを初期化した後に（ステップ625）、ディスクアレイ初期化完了通知（ステップ630）を受信待ちし（ステップ600）、受信するとNASメモリ1452の初期化を実行する（ステップ605）。

【0055】

次に、初期化が完了したことをNAS制御ボード管理プログラム1406に通知する（ステップ610）。NAS制御ボード管理プログラム1406は、その通知受信すると（ステップ635）、コマンド変換プログラム1453をディスクドライブ150からロードし、NASメモリ1452に転送する（ステップ640）。転送には、DMA等の手段を用いる。

【0056】

次に、コマンド変換プログラム1453の転送完了を、コマンド変換プログラム1453の起動プログラムに通知する（ステップ645）。コマンド変換プログラム1453の起動プログラムは、コマンド変換プログラム1453のエントリポイントへジャンプし（ステップ620）、コマンド変換プログラム1453の実行を開始する。

【0057】

次に、図7ないし図9によりディスクアレイ制御プログラムがNAS制御ボードを停止させるときの処理について説明する。

図7は、ディスクアレイ制御プログラムがNAS制御ボードを停止させるときの処理の流れを示すフローチャートである。

図8は、図7に示したフローチャートの処理から呼ばれるサブルーチンのNAS計画停止処理の流れを示すフローチャートである。

図9は、図7に示したフローチャートの処理から呼ばれるサブルーチンのNAS強制停止処理の流れを示すフローチャートである。

【0058】

ディスクアレイ制御プログラムがSAN/NASのコマンド処理をおこなっているとき（ステップ700）、管理端末60からのNAS制御ボード145の停止指示があると（ステップ705）、NAS計画停止処理を実行する（ステップ710）。NAS計画停止処理の詳細は、後に説明する。

【0059】

また、コマンド変換プログラム1453から一定間隔で受信するハートビート信号のタイムアウト（ステップ725）により、NAS制御ボード145の異常を検出した場合、コマンドキューからのNASコマンドの破棄を実行し（ステップ730）、NAS強制停

止処理を実行する（ステップ735）。NAS強制停止処理の詳細も、後に説明する。ステップ710のNAS計画停止処理およびステップ735のNAS強制停止処理の実行後は、SANコマンドを処理するSANのみの運用に遷移する（ステップ712）。

【0060】

また、管理端末60からの指示によりNAS制御ボード145を再起動した場合、再びSAN/NASコマンド処理ステップ700に遷移する。NAS制御ボード145の再起動は、ディスクアレイ制御プログラム1405が図6のステップ630に制御を移すことで実行できる。

【0061】

図8に示される計画停止処理では、NAS制御ボード管理プログラム1406は、コマンド変換プログラム1453が、計画停止通知をおこなう（ステップ800）。コマンド変換プログラム1453は、計画停止通知を受信すると（ステップ815）、ファイルキャッシュ1454の未書き込みデータをディスクキャッシュ1402に転送し（ステップ820）、停止完了の通知を実行する（ステップ825）。

【0062】

NAS制御ボード管理プログラム1406は、停止完了通知を受信すると（ステップ805）、NAS制御ボード145の停止を記録し（ステップ810）、以降NAS制御ボード145の監視はおこなわない。

【0063】

図9に示される強制停止処理では、障害発生によりコマンド変換プログラム1453のハングアップしており（ステップ905）、無限ループに陥り（ステップ910）、一切の通信を受け付けなくなる。NAS制御ボード管理プログラム1406は、コマンド変換プログラムの強制停止をNASプロセッサ1457への停止信号等ハードウェア手段により実行する（ステップ900）。これによってNAS制御ボード145は通電を停止してもデータの喪失およびSANへの影響は無い状態となり、NAS制御ボード145の交換等が実行できる。

【0064】

本実施形態では、SAN機能を持つハードウェアであるSAN制御ボード140の運用が、NAS機能を持つハードウェアであるNAS制御ボード145に依存しない構造となっているため、NAS制御ボード145のハードウェア障害時や、コマンド変換プログラム1453のソフトウェア障害時等、NAS機能の障害時にもSAN機能の運用は影響を受けずに継続することが可能である。また、NAS機能のみの計画停止や運用再開による保守をおこなうことができる。すなわち、SAN制御ボード140上のプロセッサにより動作するディスクアレイ制御プログラム1405は、NAS制御ボード145から送信されるファイル単位のI/Oコマンドの有無の如何によらず、SAN制御ボード140がSAN50よりブロック単位のI/Oコマンドを受け付けて、動作することが可能である。

【0065】

さらに、NAS制御ボード145とSAN制御ボード140の通信に内部ネットワークや共有メモリを必要としないため低いハードウェアコストで装置を実装できる。また、NAS制御ボード145は、ディスクアレイ制御プログラム1405が動作するプロセッサが不要であるため、低いハードウェアコストで実装できる。

【0066】

しかも、図5に示したようにNASコマンド515を管理するキューと、SANコマンド565を管理するキューを別にすることで、NASまたはSANのI/Oのいずれかを優先的に処理することができる。

【0067】

また、図2に示したSANプロセッサ1409、SANメモリ1407をSAN制御ボード140上に二組以上実装し、NASコマンド515を管理するキューと、SANコマンド565を管理するキューを別のプロセッサ、メモリによって処理する実施形態もあり得る。この形態によれば、NASクライアント20からのI/OとSANクライアント3

0からのI/Oが、それぞれの処理系統で独立に処理されるため、一方の処理系統の負荷が高いときにも、影響を受けることなく、その他の処理系統の性能が保証され、システム全体の性能計画を容易におこなうことができる。

【0068】

〔実施形態2〕

以下、本発明に係る第二の実施形態を、図10および図11を用いて説明する。

【0069】

先ず、図10により本発明に係る第二の実施形態の構成について説明する。

図10は、SAN/NAS制御ボードの詳細なブロック図である。

【0070】

本発明の第二の実施形態は、第一の実施形態における図1のSAN制御ボード140およびNAS制御ボード145に相当する部分を、図10のSAN/NAS制御ボード1000で置き換えたものである。

【0071】

SAN/NAS制御ボード1000のLAN40、SAN50、管理端末60、ディスク装置105に対するインタフェースは、SAN制御ボード140およびNAS制御ボード145と同じである。第一の実施形態との違いは、SANプロセッサ1409とNASプロセッサ1457がメモリコントローラ1025およびメモリ1005を共有している点である。このメモリ1005をNASプロセッサ用領域1010とSANプロセッサ用領域1015とに分割して使用する。

【0072】

コマンド変換プログラム1453は、NASプロセッサ1457で動作し、LANコントローラ1451が受信したファイルコマンドを解釈し、論理ブロックコマンドに変換する機能を持つ。コマンド変換プログラム1453は、作成した論理ブロックコマンドをSANプロセッサ1409に送信する。作成した論理ブロックコマンドをNASプロセッサ1457からSANプロセッサ1409へ送信するとき、コマンド変換プログラム1453は、共有領域1020に論理ブロックコマンドを転送する。共有領域1020は、SANプロセッサ1409とNASプロセッサ1457がデータの送受信のために使用する。ディスクアレイ制御プログラム1405は、ポーリングまたはプロセッサ間割り込みによってSANプロセッサ1409からの論理ブロックコマンドを認識し、第一の実施形態と同様に物理ブロックコマンドに変換する機能を持つ。

【0073】

SAN/NAS制御ボード1000上のNASメモリコントローラ1025以外のデバイスは、SANプロセッサ1409とNASプロセッサ1457で共有しない。ファイバチャネルコントローラ1401、ディスクキャッシュ1402、ディスクコントローラ1403は、SANプロセッサ1409が、また、LANコントローラ1451は、NASプロセッサ1457が占有して使用する。メモリコントローラ1025に内蔵されている割り込みコントローラのルーティング設定により、各デバイスからの割り込みは、それらを占有して使用するプロセッサの割り込みを発生させる。

【0074】

次に、図11を用いて本発明の第二の実施形態に係るコマンド変換プログラムとディスクアレイ制御プログラムの起動処理について説明する。

図11は、本発明の第二の実施形態に係るコマンド変換プログラムとディスクアレイ制御プログラムの起動するときの処理の流れを示すフローチャートである。

【0075】

本実施形態では、SANプロセッサ1409とNASプロセッサ1457がメモリコントローラ1025を共有しているので、コマンド変換プログラム1453のみをハードウェア的に起動・停止させることができない。したがって、コマンド変換プログラム1453を起動または再起動させる場合、先ず、NAS制御ボード管理プログラムは、NASプロセッサ用領域1010のゼロクリアを実行する（ステップ1100）。次に、コマンド

変換プログラム 1453 のディスクから NAS プロセッサ用領域 1010 へのコピーを実行し (ステップ 1105)、SAN プロセッサ 1409 への完了通知をおこなう (ステップ 1110)。

【0076】

NAS プロセッサ 1457 は、マルチプロセッサシステム固有の他系起動指示用プロセッサ割り込みの受信待ちをしており (ステップ 1115)、割り込み受信後は、コマンド変換プログラム 1453 のエントリポイントへのジャンプを実行し (ステップ 1120)、同プログラムの実行を開始する。

【0077】

本実施形態は、第一の実施形態と比較して部品点数が少なく、より低コストでの実現が可能である。第一の実施形態は、メモリコントローラおよび付随する周辺回路が二組以上必要であるのに対して、本実施形態は、一組で済む。また、プロセッサ間通信に、両プロセッサが共有して使用するメモリ内の共有領域 1020 を使用するため、図 2 に示したプロセッサ間通信コントローラ 1408 のようなハードウェアも必要としない。一方で、本実施形態では、第一の実施形態のように、SAN 機能の運用に影響を与えずに、NAS 用ハードウェアの通電中の交換はできない。

【図面の簡単な説明】

【0078】

【図 1】本発明の第一の実施形態に係る記憶装置を用いた計算機システムの構成図である。

【図 2】NAS 制御ボードと SAN 制御ボードの詳細なブロック図である。

【図 3】コマンド変換処理の流れを示すフローチャートである。

【図 4】コマンドのキューのデータ構造の例を示す図である (その一)。

【図 5】コマンドのキューのデータ構造の例を示す図である (その二)。

【図 6】本発明の第一の実施形態に係るコマンド変換プログラムとディスクアレイ制御プログラムの起動するときの処理の流れを示すフローチャートである。

【図 7】ディスクアレイ制御プログラムが NAS 制御ボードを停止させるときの処理の流れを示すフローチャートである。

【図 8】図 7 に示したフローチャートの処理から呼ばれるサブルーチンの NAS 計画停止処理の流れを示すフローチャートである。

【図 9】図 7 に示したフローチャートの処理から呼ばれるサブルーチンの NAS 強制停止処理の流れを示すフローチャートである。

【図 10】SAN/NAS 制御ボードの詳細なブロック図である。

【図 11】本発明の第二の実施形態に係るコマンド変換プログラムとディスクアレイ制御プログラムの起動するときの処理の流れを示すフローチャートである。

【符号の説明】

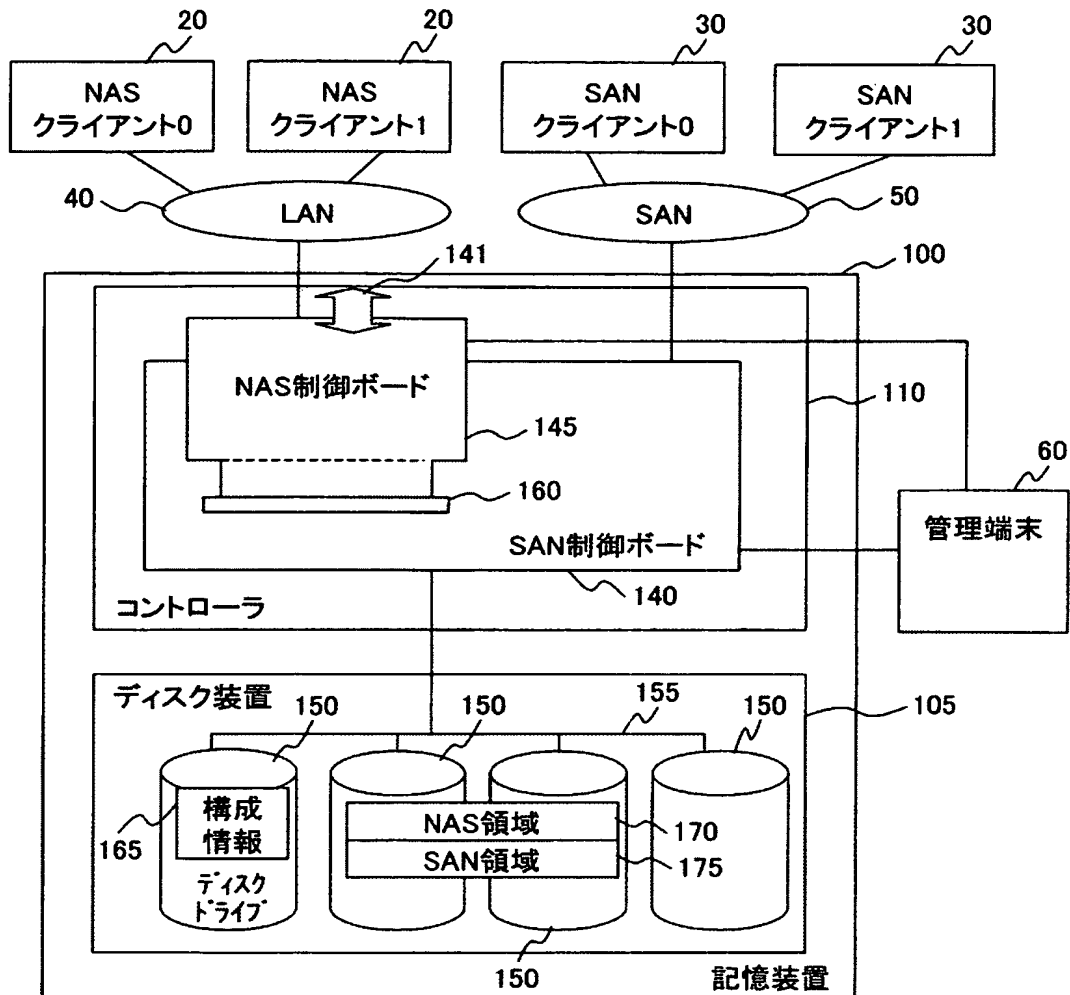
【0079】

20...NAS クライアント、30...SAN クライアント、40...LAN、50...SAN、60...管理端末、100...記憶装置、105...ディスク筐体、110...コントローラ筐体、140...SAN 制御ボード、145...NAS 制御ボード、150...ディスク、155...通信路、160...コネクタ、1401...ファイバチャネルコントローラ、1402...ディスクキャッシュ、1403...ディスクキャッシュ、1404...通信用領域、1405...ディスクアレイ制御プログラム、1406...NAS 制御ボード管理プログラム、1407...SAN メモリ、1408...プロセッサ間通信コントローラ、1409...SAN プロセッサ、1410...SAN メモリコントローラ、1451...LAN コントローラ、1452...NAS メモリ、1453...コマンド変換プログラム、1454...ファイルキャッシュ、1455...通信用領域、1456...NAS メモリコントローラ、1457...NAS プロセッサ、1000...SAN/NAS 制御ボード、1005...NAS メモリ、1010...NAS プロセッサ用領域、1015...SAN プロセッサ用領域、1020...共有領域。

【書類名】 図面

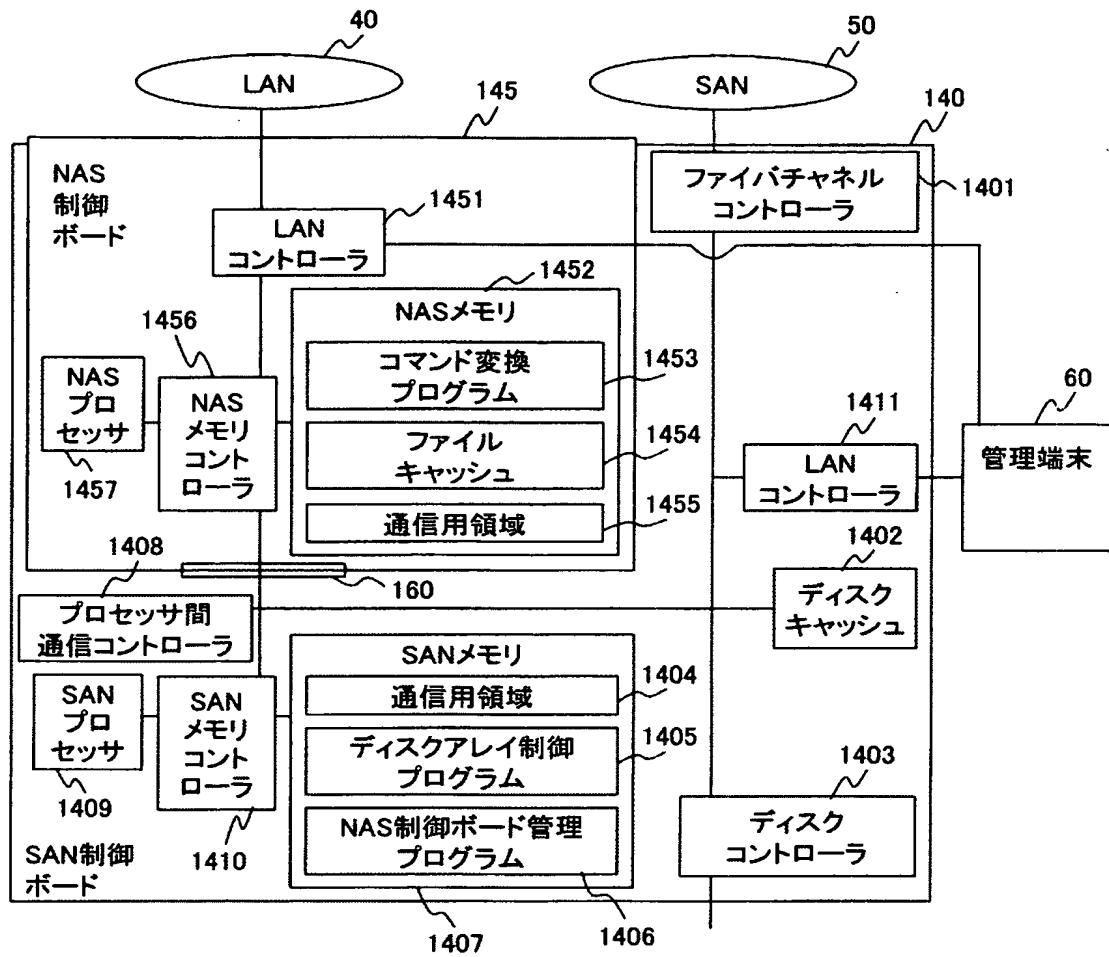
【図 1】

図 1



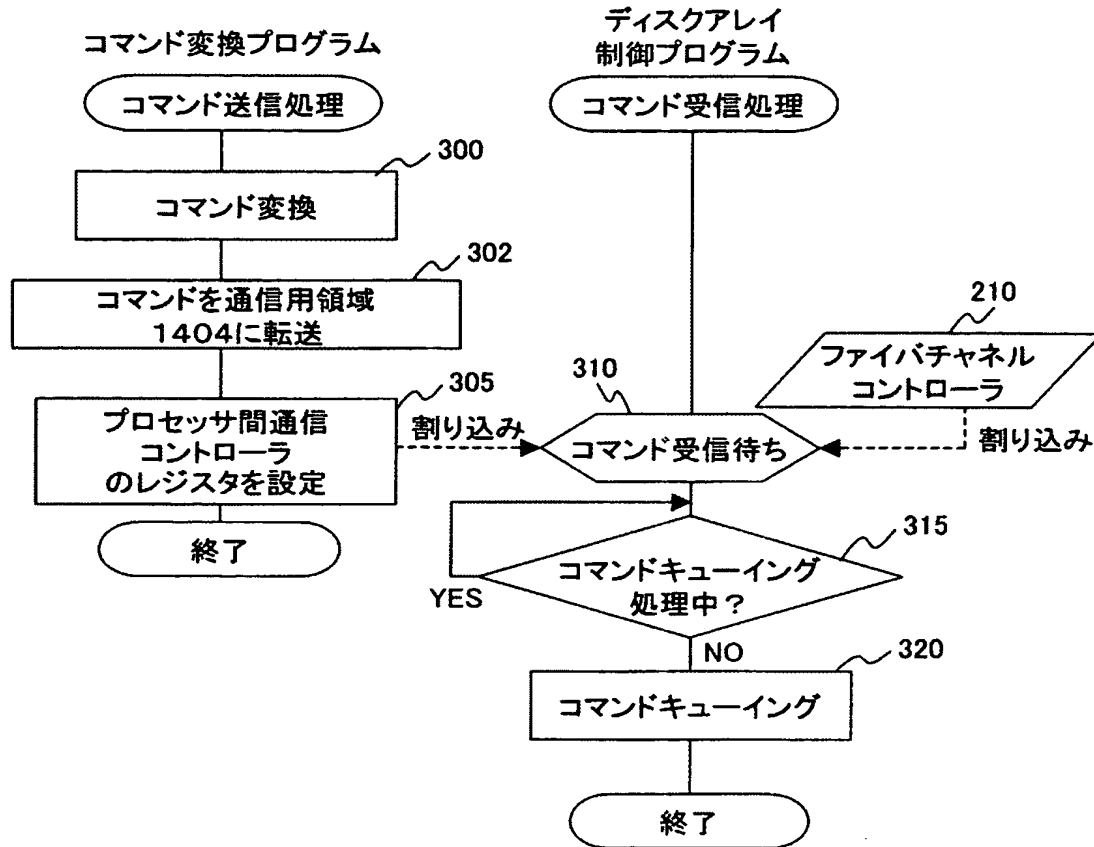
【図 2】

図 2



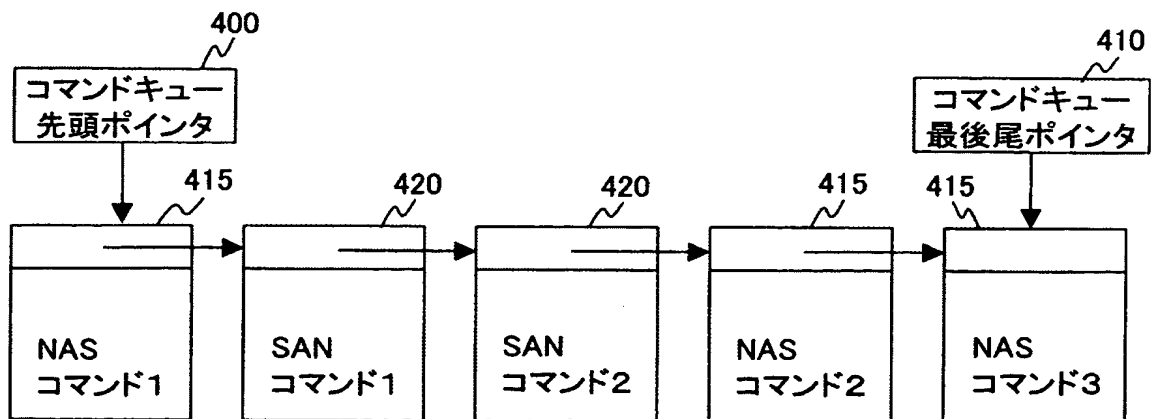
【図 3】

図 3



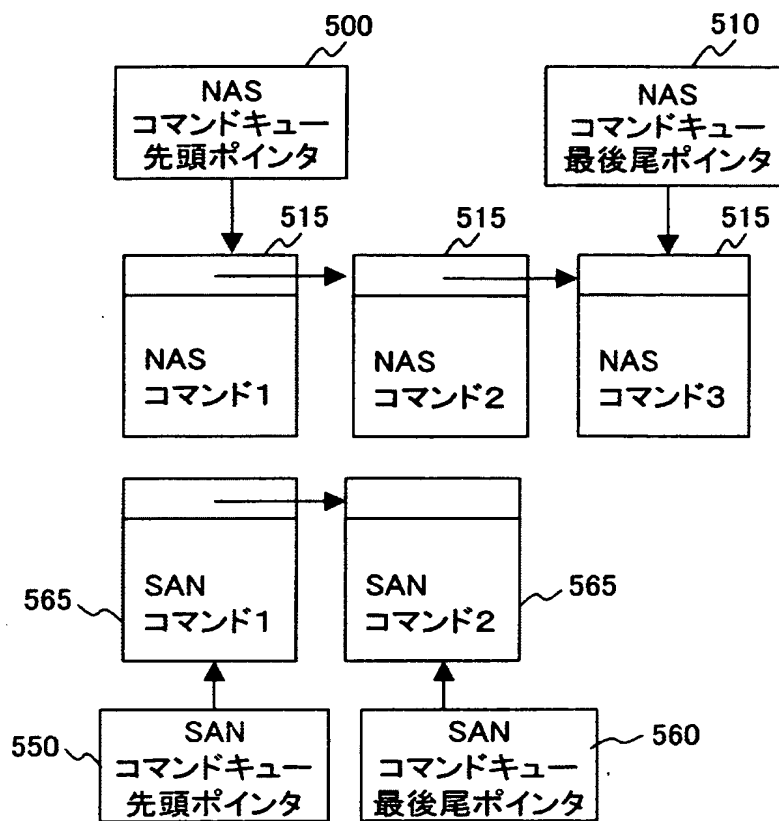
【図 4】

図 4



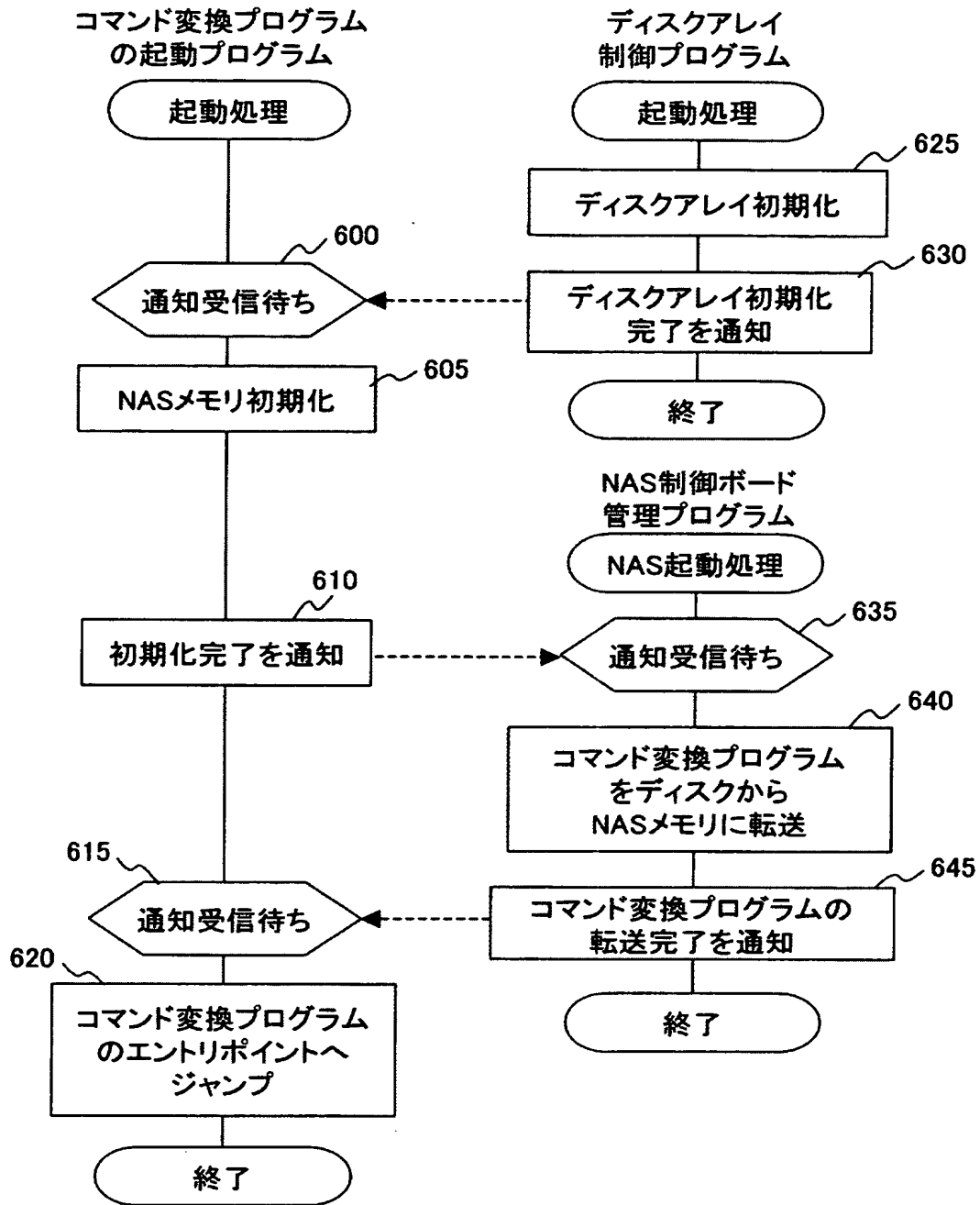
【図 5】

図 5



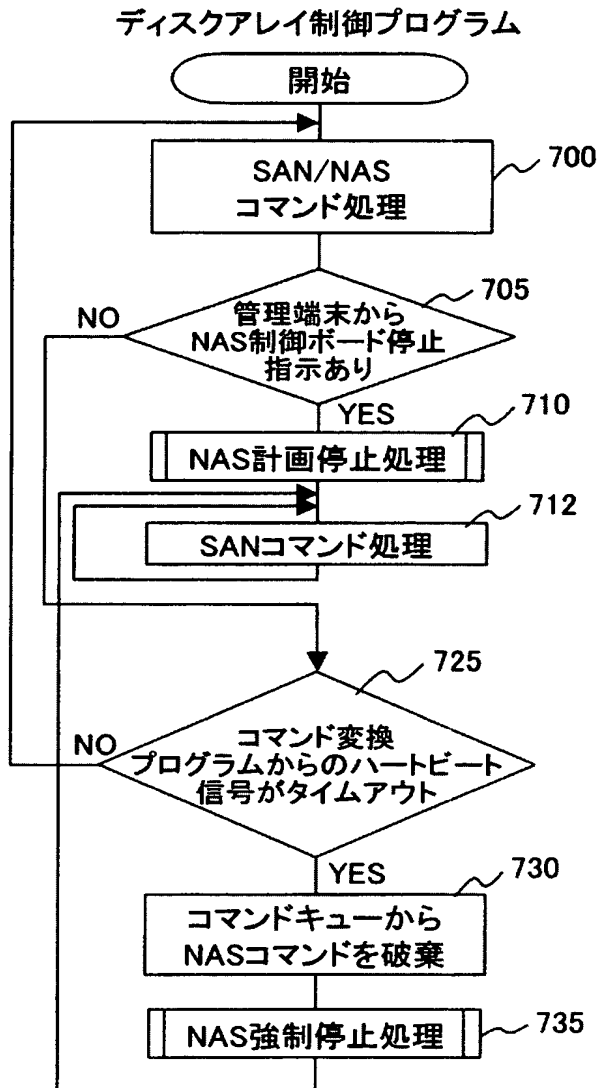
【図 6】

図 6



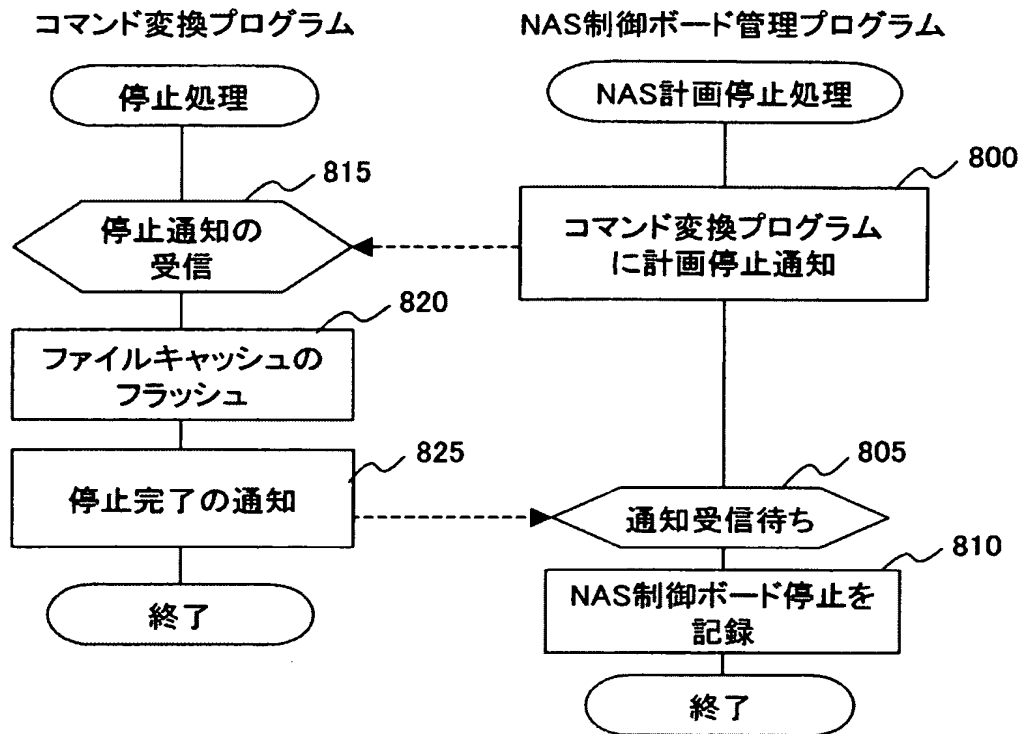
【図 7】

図 7



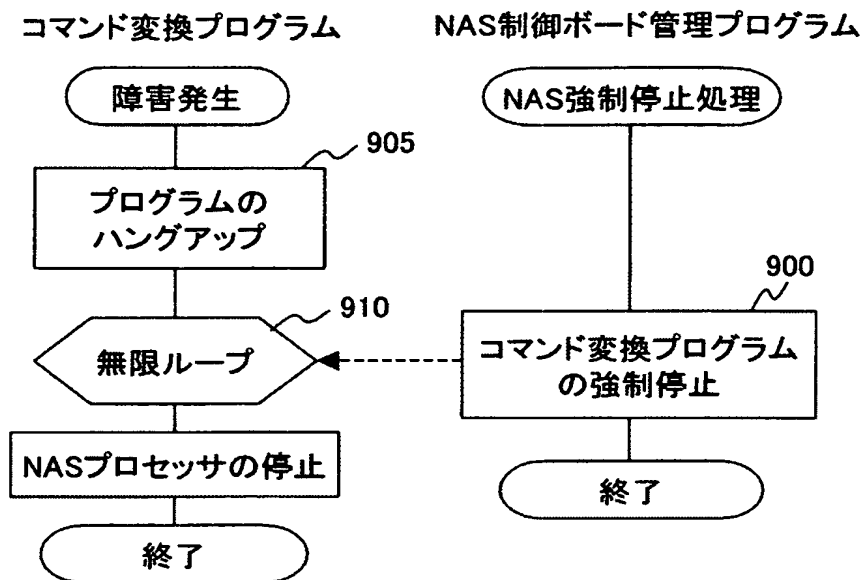
【図 8】

図 8



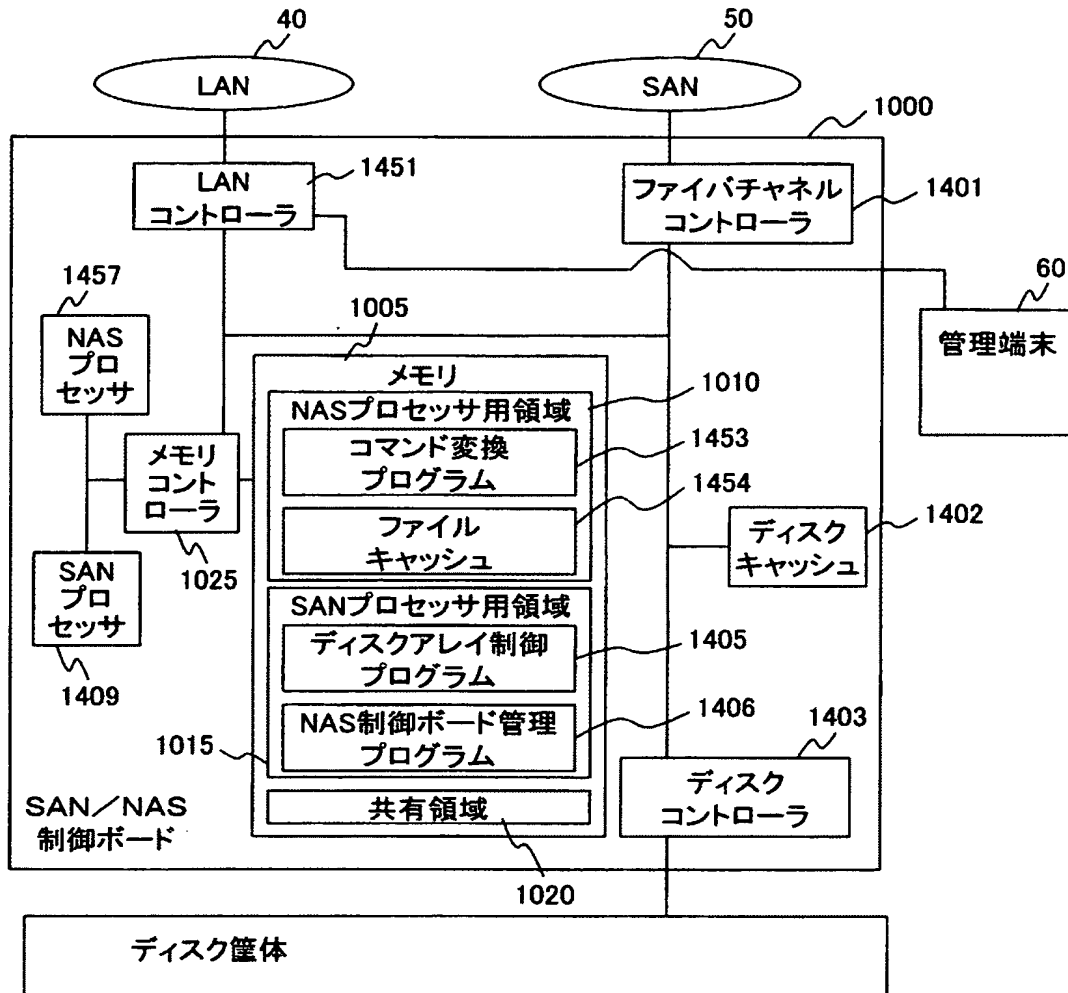
【図 9】

図 9



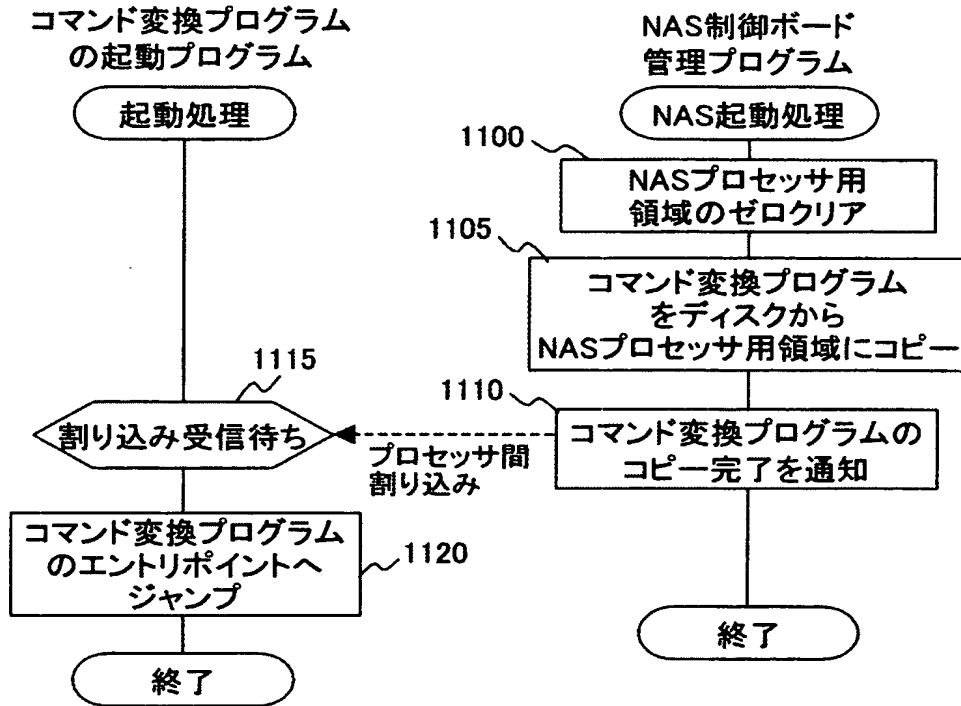
【図 10】

図 10



【図 11】

図 11



【書類名】 要約書**【要約】**

【課題】 S A N インタフェースと N A S インタフェースの両者を有するハイブリッド型の記憶装置において、N A S 機能の障害や N A S の負荷などにかかわらず、S A N 機能として独立に動作できるようにし、しかも、簡易なハードウェアで実現する。

【解決手段】 記憶装置のコントローラを、ファイル単位の I / O コマンドを受け付ける N A S 制御ボードと、ブロック単位の I / O コマンドを受け付ける S A N 制御ボードとで、独立に動作できるように実装する。そして、N A S 制御ボードが、N A S インタフェースのファイル単位の I / O コマンドを、ブロック単位の I / O コマンドに変換して、S A N 制御ボードに送信する。S A N 制御ボードは、S A N から受け付けたブロック単位の I / O コマンドと、N A S 制御ボードから受信したブロック単位の I / O コマンドとによりディスク装置に記憶されたデータをアクセスする。

【選択図】 図 2



特願 2 0 0 4 - 0 2 1 2 0 8

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 1 0 8]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所